Bibliometrics Working Group

Co-chairs: Sarah Callaghan - presenter (sarah.callaghan@stfc.ac.uk)
Kerstin Lehnert (lehner@LDEO.COLUMBIA.EDU)
What are Bibliometrics?

- Quantitative measures to assess and measure impact and quality of research and researchers,
- by tracking and recording access and citation of scientific publication.
- Inform & influence the advancement of academic careers, investments in research and scientific journals.

The basic idea of bibliometrics is to evaluate the attention scientific publications receive within the scientific community.
Bibliometrics: Examples

- ISI Science Citation Index – *individual articles*
- h-index (= Hirsch number) – *researchers*
- Impact Factor - *journals*

Bibliometrics for Data

• Quantitative measures for the use, utility, and impact of data.

• **Rationale**: Essential for a culture change toward full appreciation and recognition of data as a part of the scholarly record.
  
  • raise the value of data acquisition, curation, and sharing;
  • encourage more and better data citation;
  • augment the overall availability and quality of research data;
Use of existing metrics?

- Data citation is not standard practice in the research community - though it is getting more common
- Though Principles for Data Citation (Joint Declaration on Data Citation Principles, CODATA TG on DataCitation) exist, work is still being done on how to implement them (Force11)
- Data repositories are mostly not ready for the data publication concept and have not implemented formal procedures.
- There are unresolved issues of granularity, versioning, unique identification, metadata, review.
Goals

• “conceptualize data metrics and corresponding services to overcome barriers”
  - Summarize current and emerging *data citation practices* of all stakeholders (journals, data centers, funders, societies);
  - Understand and articulate *necessary organizational and cultural changes* in the scholarly publishing system needed to foster proper attribution of data sets;
  - Evaluate and report on *possible models* how data use and citations can be successfully tracked and measured based on existing and emerging approaches (e.g. altmetrics);
  - Identify and report on *possible barriers* for the implementation and adoption of data citation and data metrics solutions.
Members

Note that the membership of this WG is open, and members are actively seeking participation from members outside the US/EU

- Kerstin Lehnert (US, IEDA, WDS) [CO-CHAIR]
- Sarah Callaghan (UK, BADC) [CO-CHAIR - also co-chair of the CODATA TG on data citation]
- Jan Brase (Germany, DataCite) [also co-chair of the CODATA TG on data citation]
- Ross Cameron (The Netherlands, Scopus)
- Cyndy Chandler (US, Woods Hole Oceanographic Institution)
- Ingeborg Meijer (The Netherlands, University of Leiden)
- Fiona Murphy (UK, Wiley-Blackwell)
- Lyubomir Penev (Bulgaria, Pensoft Publishers)
- Fiona Nielsen (UK, DNAdigest.org)
- Nigel Robinson (UK, Thomson Reuters)
- Mary Vardigan (USA, ICPSR)
- Jochen Schirrwagen (Germany, Universität Bielefeld)
Case Statement

• Outlines rationale, requirements, benefits, challenges, approaches, e.g.
  • Bibliometrics are essential for a culture change.
  • Will allow stakeholders to formally & quantitatively demonstrate significance and viability of data for science.
  • All stakeholders need to embrace the system as credible, valuable, and meaningful.
  • Engage with existing initiatives and efforts
    • DataCite, TR DCI, ICSU CODATA WG Data Citation/Force 11, CrossRef, Altmetrics, Mendeley, ImpactStory, Scopus;
• Presents the workplan and plan for adoption.
Case Statement

• Workplan
  • Evaluate possible approaches - via a survey (impact & feasibility)
    • Altmetrics, Data Citation Index, Mendeley, F1000R with figshare, Elsevier’s linking with some data centers.
    • future possibilities with CrossRef, DataCite, Thomson Reuters Industry Forum, STM Association, EarthCube, and others.

• Survey task breakdown
  • Determine audience: which questions to target which groups?
  • Develop questions to evaluate current approaches outlined in our summary.
  • Develop questions to ask for feedback for our user requirements —“Do the current practices meet your needs?”, “What are your bibliometrics needs?”, “If you are aware of bibliometric tools, but do not use them, why? And what would encourage you to use them?”— and needs for information/communication of bibliometrics tools. The questions will be a combination of ranking on a scale and free text, as appropriate.
  • Keep the survey short, and phrase questions and preamble to encourage responses from broad range of stakeholders.

• Develop recommendations
  • Organizational, technical, financial, methodological
Recommendations

• What are the building blocks for an optimal system?
• What changes are needed from funders, policy makers, data centers, science publishers, learned societies, academic institutions to implement data citation & metrics?
• Do we need commonly operated services?
• What technical components, interfaces, standards are needed? What is currently available, usable & appropriate?
• How can necessary technical requirements be implemented most effectively & efficiently?
• How community-specific do recommendations need to be to be adopted & implemented?
• What are the costs for data metrics?
Case Statement

• Adoption Plan
  • Recommendations for all stakeholders that will include
    • Case studies, practical next steps
    • Examples for variety of bibliometric types, data types, domains
    • General requirements for citability of data
  • Broader community engagement and participation
    • webinars, intermediate reports, etc.
• Communication plan
  • Regular teleconferences (min. 6 weeks) with sWG and with WG
  • Publishing Data
  • Form subgroups to take on specific sections and projects
  • Workshop Nov 4-6, 2013 in Gran Canaria
  • Open webinar for plenary exchange of the first results
Deliverables

• Case studies: Based on the results of the survey and existing experiences of the WG team.
  • Identify potential partners by name/organization. Ideally involve them and achieve buy-in from these entities ahead of publication of the deliverable itself to ensure forward momentum.
  • Clear list of bibliometric examples and potential use cases. Include a variety of data types, subject areas and bibliometric types (citation, usage, social media as appropriate)
  • Examples of situations where bibliometrics for data have impacted researchers’ professional lives - to encourage and promote bibliometrics for data uptake by data users and producers
• General requirements for citability of scientific data (granularity, citation information and persistent identification) - rely on CODATA TG on data citation report and the response to survey (described in work plan above) where possible.
• Use cases and requirements: to provide guidance on concrete, practical next steps. Plans for implementation of a bibliometrics system (as input to the Publishing Services WG) and user consultation with stakeholders on those plans.
Time plan

<table>
<thead>
<tr>
<th>2014Q1</th>
<th>2014Q2</th>
<th>2014Q3</th>
<th>2014Q4</th>
<th>2015Q1</th>
<th>2015Q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey and evaluation</td>
<td></td>
<td></td>
<td>User feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop recommendations</td>
<td></td>
<td></td>
<td>Case studies and examples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requirements and use cases for implementation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M1: survey launch
M2: 4th RDA – preliminary survey results
M3: survey close
M4: Final recommendations
Feedback

• Are we missing relevant stakeholders?
• Are we missing relevant initiatives to engage with?
• Are we focusing on the essential aspects of bibliometrics and data citation?
• How can we improve our approach and outcomes?

• Please join us!

• Contact: sarah.callaghan@stfc.ac.uk, lehnert@LDEO.COLUMBIA.EDU